RL-TR-91-48
In-House Report
April 1991

MEASURING THE QUALITY OF KNOWLEDGE WORK

Anthony Coppola

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

Rome Laboratory
Air Force Systems Command
Griffiss Air Force Base, NY 13441-5700

This report has been reviewed by the Rome Laboratory Public Affairs Office (PA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

RL-TR-91-48 has been reviewed and is approved for publication.

APPROVED:

ANTHONY J. FEDUCCIA, Chief
Systems Reliability and Engineering Division
Directorate of Reliability and Compatibility

APPROVED:

JOHN J. BART
Technical Director
Directorate of Reliability and Compatibility

FOR THE COMMANDER:

RONALD RAPOSO
Directorate of Plans and Programs

If your address has changed or if you wish to be removed from the Rome Laboratory mailing list, or if the addressee is no longer employed by your organization, please notify RL(RBE-1) Griffiss AFB NY 13441-5700. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. AGENCY USE ONLY (Leave Blank)	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	April 1991	In-House

4. TITLE AND SUBTITLE
MEASURING THE QUALITY OF KNOWLEDGE WORK

5. FUNDING NUMBERS
PE - 62702F
PR - 2338
TA - 02
WU - TK

6. AUTHOR(S)
Anthony Coppola

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Rome Laboratory (RBE-1)
Griffiss AFB NY 13441-5700

8. PERFORMING ORGANIZATION REPORT NUMBER
RL-TR-91-48

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Rome Laboratory (RBE-1)
Griffiss AFB NY 13441-5700

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
N/A

11. SUPPLEMENTARY NOTES
Rome Laboratory Project Engineer: Anthony Coppola/RBE-1/(315) 330-4758

12a. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE

13. ABSTRACT (Maximum 200 words)

There are a variety of ways in which the quality of knowledge work can be measured, depending on the definition of quality and the intended use of the measure. This report summarizes these for the guidance of the managers of knowledge workers, such as the engineers and scientists of Government laboratories.

14. SUBJECT TERMS
Quality, Quality Measurement, Total Quality Management

15. NUMBER OF PAGES
28

16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

Foreword

This report discusses the measurement of quality in organizations performing knowledge work. It will attempt to clarify both the difficulties and the possibilities. Since there is little on this in the literature, this work is essentially the perspective of the author.

This report is dedicated to Colonel Raymond A. Shulstad who issued the challenge to produce it.

Table of Contents

1. Introduction

The word "quality" gets a lot of use these days. Manufacturers of hard goods have recognized that producing quality products is their main hope for survival in the face of fierce competition. Hence, the extensive use of the word in their advertisements, often accompanied by some real improvements in their products. Where real improvements are made, you will find organized and sustained quality efforts based on a set of effective principles. Understanding the principles for achieving quality is also of vital interest to the managers of knowledge workers such as the engineers and scientists of government research and development laboratories. These agencies are now in fierce competition for a shrinking pot of available funds. Like the hardware producers, laboratories producing quality products have the best, perhaps only, chances for survival.

Quality education has become a growth industry, with a multitude of "gurus" available to meet the demand for guidance. While these experts do not agree on everything, they have some common tenets. One of these is that quality must be measured if it is to be improved. Without measurement, an effort to improve quality may be full of sound and fury but in the end will change nothing. And therein lies the rub for laboratory managers. How do you measure the quality of knowledge work?

It is the objective of this report to answer that question.

Let's start by looking at current practice. In the research for this report, despite occasional declarations of impossibility, I found a variety of quality measures currently in use by laboratories, including:

- Customer ratings
- The number of patents, papers or advanced degrees among the people
- Measures of the "climate" in the laboratory, such as absentee rate
- Adherence to budget and schedule
- Test results on hardware and software products
- The amount of external funding
- Contracting cycle time

Which, if any, is best? As usual, it depends. First of all, it depends on how we define quality. It also depends on our reasons for measuring quality. Let's take these one at a time.

There is no standard definition for quality. Indeed, there are so many definitions of quality, that it makes more sense to examine them by category.

David A. Garvin (1) identifies five categories of definitions for quality. These are:

1. Transcendent: a subjective feeling of "goodness".

2. Product-Based: measured by attributes of the product

3. Manufacturing-based: conformance to the specifications

4. Value-based: "goodness" for the price

5. User-Based: the capacity to satisfy the customer

Each of these categories stems from definitions coined by analysts attempting to meet their particular quality needs. We should note that the categories are not mutually exclusive. In particular, no matter what definition is used, ultimately quality is always defined by the customer (i.e. user-based). If an agency feels its quality is excellent (gives itself a high a transcendent quality rating), and its customers think otherwise, the agency may confidently continue practices which lead to its destruction. Similarly, if quality is measured by attributes or conformance to specifications, and the attributes or requirements selected do not reflect the voice of the customer, the analyst is deluding himself. Finally, value-based measures must reflect the value perceived by the customer, or the product may share the fate of the Edsel. Thus all roads to defining quality lead to the customer, or they go nowhere.

Any quality definition used must be compatible with the other concern mentioned, the purpose of the measurement. Within the umbrella of measuring quality, we could be attempting to gauge customer satisfaction, appraise the agency's overall quality, appraise an individual's performance, or improve specific products, services and processes. This report's objective can therefore be restated as filling in the blanks on the following matrix:

MEASURES OF KNOWLEDGE WORK

PURPOSE:	RATE CUSTOMER SATISFACTION	APPRAISE AGENCY	APPRAISE INDIVIDUALS	IMPROVE PRODUCTS AND PROCESSES
MEASURE:				
TYPE OF MEASURE:				

In the following chapters we will examine each category in Garvin's Taxonomy, identifying measures appropriate to knowledge work, noting their advantages and drawbacks, and matrixing the measures against measurement objectives. After all five categories have been covered, we will combine the results and discuss our findings based on a consideration of laboratory priorities.

2. Transcendent Quality Measures

"...Even though Quality cannot be defined, You know what it is," said Robert M. Pirsig in "Zen and the Art of Motorcycle Maintenance." Pirsig's statement epitomizes the theory behind transcendent quality measures, which are merely means for capturing subjective opinions.

The most common tool for transcendent quality measurement is the rating scale. For example, cake mixes are tested by submitting their products to a panel who rate the taste of the cake on a scale from one to five, with five being the best possible. Knowledge workers sometimes use peer ratings in a similar manner. Currently, all agencies in the Air Force Systems Command are developing customer surveys to obtain transcendent quality ratings.

When an attribute is actually subjective, like taste, the transcendent cannot be challenged. In areas where other measures are possible, the more objective measures are generally preferable. Even then, when practical difficulties prevent the use of better measures, subjective opinion may be useful, so long as it reflects the opinion of the customer. In fact, the transcendent opinion of the customer is the most important measure of one's quality.

A danger to avoid is using the producer's opinion instead of the customer's. Surveys have shown that executives universally consider the quality of their agencies better than average. They can't all be right, and the complacency brought about by this belief can easily become the foundation of a disaster. There is an illustrative story of a Japanese failure (yes, they have them too), caused by an incorrect self-evaluation. A Japanese candy manufacturer advancing in years made his own taste test of a proposed new product and decided it was good enough to market. Unfortunately, his much younger customers had different tastes and the product did not sell.

In my opinion, a useful area for transcendent measures of quality is in individual performance appraisals.

Dr. W. Edwards Deming, the most respected "guru" of quality, condemns the use of annual appraisals for several reasons. (2) He points out that they encourage short term performance over long term, and individual performance over teamwork, both of which are destructive to the agency involved. Also, he notes that appraisals seldom account for normal variation in a process. In any process, most results will be distributed about an average value. Half will always be below average, by definition. An average worker will produce below average results half the time. Hence, his appraisal can become a lottery, with his reward or lack thereof determined by chance.

Despite Dr. Deming's condemnation, appraisal systems will probably be with us for a while. The use of transcendent measures may be one way to make them work. My recommendation is to use general categories (e.g. shows initiative) rather than specific (e.g. supplies five ideas for new projects annually), scored by the subjective opinion of the employee's supervisor, on the assumption that the supervisor's transcendent quality judgement of the employee is likely to be an accurate measure (He will know quality work when he sees it). Another possibility is peer rating, which would also require radical changes to existing appraisal systems.

Alternates to performance appraisals do exist which permit the use of more objective measures. Profit sharing plans are one way to reward good work. They will also create a peer pressure for quality as each worker's performance will affect his colleagues' pocketbooks. A government laboratory could perhaps create a pool of money based on the efficiency of its operations, determined using overhead rates or cost of quality measures (discussed in Chapter 5). This would spur teamwork and greatly encourage employee challenges to non-productive management practices. Until such alternates are established, I recommend transcendent definitions of quality for individual appraisal.

Finally, even when using more objective quality definitions, the transcendent can be useful as a "sanity check". If a measured quality value "feels" too high or too low, perhaps your intuition is telling you to reevaluate your selection of measures. But be careful; don't let your ego tell you that you are better than you really are.

Transcendent definitions of quality are of no help in determining how to improve, and in measuring progress of the improvements, except in a gross sense. For these uses, other measures are much more desirable.

Summarizing the above in a matrix:

TRANSCENDENT QUALITY MEASURES OF KNOWLEDGE WORK

PURPOSE:	RATE CUSTOMER SATISFACTION	APPRAISE AGENCY	APPRAISE INDIVIDUALS
MEASURE:	Rating scales of customer opinions	Rating scales of customer or peer opinions	Rating scales of supervisor's opinions
TYPE OF MEASURE:	Subjective	Subjective	Subjective

3. Product-based Quality Measures

Product-based quality is measured by the amount of some desired ingredient or attribute. For example, the speed of a fighter plane (or of a computer). In knowledge work, one desired attribute may be innovation. The difference is, of course, that it is easy to measure speed.

Since innovation and other intangible features are desired not for themselves, but for their impact on the product, measurable units such as speed will reflect the quality of knowledge work once the work is transitioned into hardware or software. Under such circumstances, system parameters can be measured to establish the quality of the underlying knowledge work. This doesn't mean it is easy. There are many parameters of, say, an electronic system, which represent desirable attributes. Unless a few dominate, one can be swamped in measures. One can try to select the most meaningful measures, which should be the main interests of the product's user, and the main reasons the product was developed. To be effective as quality measures, however, the measured values must be referenced to some benchmarks. For example, the speed of a computer is useless for quality evaluation unless the analyst knows what previous machines delivered. Percent improvement in a parameter over previous achievements is an appropriate measure of the quality of the improvement effort.

Besides picking the critical parameters, a problem with attribute measures is that trade-offs may not be recognized. Speed may be enhanced at the expense of payload which may or may not be an improvement overall. One way to evaluate this is the use of all-encompassing measures such as "systems effectiveness." Systems effectiveness is defined as a function of a system's availability, dependability and capability against a specified threat (3). In the simplest case, availability is the probability of a system being operable when needed, dependability the probability that it will remain operable for the length of a mission and capability the conditional probability that, if operating, it will successfully complete the mission. For this simple case:

System Effectiveness = (Availability)*(Dependability)*(Capability)

When one begins to consider degraded mission states, variations in the threat, ability to repair, etc., this simple formula expands to a problem in matrix algebra. Those wishing to pursue it further are directed to reference 3.

An approach between the measurement of a few selected parameters and the calculation of system effectiveness is the use of indexes. Indexes are artificial, but supposedly not arbitrary, groupings of measures into an overall single measure. Examples are the consumer price index and the index of leading economic indicators. Similarly, a quality index can be created by identifying parameters of interest, establishing measures, weighing the measures and combining them into one. As a simple example, Robert Gunning (4) describes a "fog index" for evaluating understandability of text. It is calculated by computing the average sentence length, adding this to the number of words of three syllables or more in 100 words, then multiplying by 0.4. Though Gunning claims his index corresponds roughly with the number of years of schooling a person would require to read the text with ease and understanding, an index figure is generally not meaningful in absolute terms. Rather, it shows trends, which is generally satisfactory. The results can be compared to benchmarks and can also be plotted on a control chart. Against these advantages, it is an artificial figure. If its components are not chosen carefully, it can also be an arbitrary number not particularly good as a measure of quality. Weighting can be an interesting problem. In the example, suppose we used the average number of words with three or more syllables in 50 words, rather than 100. Would we have a better or worse measure?

Indexes do not have to be limited to simple linear relationships. The technical report AFHRL-86-64 (Reference 5) provides a sophisticated indexing approach where the weight can be changed as a function of the indicator's value. It also provides a means of cascading measures, so one department's index can be combined with others to create an index of the grouped departments. Readers wanting to use indexes should obtain a copy of reference 5.

Should you use an index? When a single parameter measurement is inadequate or conflicting goals exist, an index may be a useful tool. Whether a particular index is well constructed is another question. The customer's input would be invaluable in creating a good index.

Summarizing thus far, when knowledge work is transitioned into tangible products, the parameters of the products can be used as a measure of the quality of the knowledge work applied. Measures can be single parameters (e.g. speed), overall measures of systems effectiveness, or thoughtfully constructed indexes.

Obviously, the more tangible the product, the better product-based measures work. However, in knowledge work the product is often intangible, such as a conceptual design or a set of recommendations, and product parameters cannot be measured as reflections of quality attributes. One way out is to use even more indirect measures so long as they also correlate with the the attributes desired. For example, a large number of patents should indicate an innovative agency. Although this does not guarantee that any particular product of that agency will be produced with a high degree of innovation, it can provide a "warm fuzzy feeling" to a potential customer and to the laboratory commander. Again, benchmarks are needed for accurate interpretation.

Some sample measures might be the ratio of in-house to contracted work, numbers of papers published, patents awarded, dollars spent on education and training activities, advanced degrees earned, name requests for consulting committees received, and the amount of national/international professional activity among the knowledge workers. These are measures of the laboratory climate or environment favoring quality knowledge work.

One could also measure the climate opposing quality in knowledge work. Common measures indirectly showing unfavorable climates include absenteeism, turnover, average sick days per employee, etc. Poor environments could perhaps be more directly measured by the number of approvals required to do something, the ratio of overhead to productive activity, the length of time required to obtain a part or a piece of test equipment, etc. These could be labelled "Hassle indexes."

In summary, product-based quality measures are most useful when tangible products are available. Attributes like the ability to innovate cannot be measured directly. Instead, "by their fruits ye shall know them". Measures of environment, rather than of specific products, can be used when no tangible product is available. Benchmarks are needed to evaluate the measures.

Putting this into a matrix:

PRODUCT-BASED QUALITY MEASURES OF KNOWLEDGE WORK

PURPOSE:	RATE CUSTOMER SATISFACTION	APPRAISE AGENCY
MEASURE:	Product parameters, performance indexes, system effectiveness (against benchmarks)	climate indicators - favorable signs, - "hassle indexes" (against benchmarks)
TYPE OF MEASURE:	Objective	Surrogate

4. Manufacturing-based Quality Measures

Perhaps the best illustration of manufacturing-based quality definitions was proposed by Philip Crosby, who equated quality to compliance with specifications (6). This, of course presumes tangible products or services, which for knowledge work could include such things as technical reports and briefings as well as the more obvious hardware and software end products.

The most commonly used manufacturing-based quality measure is defect rate (i.e. the percent of the product not in compliance to specifications). Defect rate is a universal quality measure and can be applied to knowledge work as well as manufacturing, though not as easily. In using defect rates, one must have an operating definition of defect. Is a misspelling a defect? Would it be considered the same in a sales brochure, a technical report, and a telegram authorizing a purchase? A reasonable operating definition must be formulated describing defects to be monitored.

Besides percent defects, there are other manufacturing-based measures of quality of varying utility to knowledge work. For example, yield is a common measure of product quality. It is simply the percent of manufactured products which are not defective. Although we could probably invent some way to apply it, it really isn't too useful in measuring knowledge work. On the other hand, cycle time is another widely used measure which is easily applied to knowledge work.

Product-based measures become manufacturing-based measures when acceptable limits are defined. For example, Gunning's "fog index," discussed in Chapter 3, can be used to specify a required value of understandability, which can then be evaluated by a manufacturing-based quality measure (e.g. percent of reports exceeding a specified "fog index").

Another manufacturing-based quality measure is the variation among products. All products will have some variation, and the greater this is, the more defects we will have. For illustration, suppose we did specify that all reports to a particular customer have a fog index no higher than 12. If our measurements show the average fog index of our reports to be 11.0, we are not necessarily doing well. We could be producing reports with fog indexes between 10 and 12, or between 9 and 13, or between 8 and 14, etc. The greater the variance, the more products out of specification, and the less predictable the quality of a single product. Variance can be measured in various ways, such as by range (the difference between the highest and lowest values) or by standard deviation (a statistical measure).

Standard deviation is estimated by taking a sample of the product and measuring each item in the sample for the value of the parameter of interest. The standard deviation of the product from which the sample came is then calculated by:

$$\text{sigma} = \sqrt{\frac{\sum_{i=1}^{n} (X_i - \bar{X})^2}{n-1}}$$

sigma = standard deviation of population sampled

X_i = unit values

$\bar{X}$ = mean value of units in sample

n = number of units in sample

Assuming a normal or bell-shaped distribution of the parameter, 99.7% of the product will have values no more than three sigmas away from the mean value. The lower the value of sigma, the more uniformity in the product.

Variance, however, cannot be the whole story. Suppose, for example, the mean fog index of our reports was 14.0 and three sigmas equaled 0.2. The understandability of our reports is quite predictable, but that would be of no comfort to the customer who needs a fog index of 12 or less. Hence, both the mean and variance are important. A measure which considers both is called process capability (Cp). It compares the mean and variance of a product parameter to specified limits.

$$\text{Cp} = \frac{\text{(upper specification limit - lower specification limit)}}{\text{6 sigma}}$$

Thus a Cp of 1.0 means that 99.7% of the product would be "in spec" assuming the mean of the product is centered between the upper and lower control limits. To allow for means in other locations, a Process Performance (Cpk) Index can be used.

$$\text{Cpk} = \frac{\text{(minimum distance between the mean and either control limit)}}{\text{3 sigma}}$$

Using either measure, the higher the value, the better. Motorola's "six sigma" program strives for a Cp of 2.0 (six sigmas between the target mean and the specification limits) which, when the true mean is 1.5 sigmas off target, translates to a defect rate of 3.4 parts per million.

For non-structured work, the main problem with manufacturing-based quality definitions is determining what the "specification" is. A specification for a study on Computer Technology may specify the format, perhaps even the type style, of the final report, which are all of secondary importance to a host of considerations such as responsiveness, innovation, realism, clarity, etc. With the exception of the fog index for understandability, I have found no specifiable measures of these critical desires. Though the ease of measuring against a specified value is seductive, it can lead to such things as the seriously made proposal that the standard for judging the performance of travel duty be how often the traveller submitted a trip report in a specified five-days. (So if you went to higher headquarters and made a perfect fool of yourself, but reported it in less than five days, would you be a hero?).

If you assume meeting the specifications for a product reflects desired intangibles like innovation, measuring conformance is adequate. Otherwise, the manufacturing-based measures simply will not work. One could, I suppose, specify that a product show innovation, but verification of compliance would require a subjective opinion, which is a transcendent, not a manufacturing-based quality measure. (Note: requirements which cannot be objectively measured are usually barred from specifications as unenforceable.)

However, manufacturing-based quality figures do have an important place in knowledge work. A laboratory's operations include many processes and sub-processes. It is important to note that in knowledge work, as in any other, the final customer is only the last of a series. Each office involved in a process is the customer for some input and the provider of some output to another customer. Thus, even the process of creating innovations will include such processes as publishing reports, obtaining laboratory equipment, awarding contracts, etc., which can be evaluated by manufacturing-based quality measures. Improving these processes must improve the laboratory operations, even if we totally ignore intangibles like innovation. For example, shortening the time to obtain a needed instrument yields more time for performing experiments with it, which in turn can produce more innovations.

Process improvement is the heart of Total Quality Management. Improving the process can be accomplished by radical innovations or by accumulation of many small changes. Either way, it begins with an understanding of the process, and depends on the measurement of quality indicators. The process itself should tell you what to measure. If the process is proposal evaluation, for example, cycle times and/or the number of corrections required (defects) may be compiled to establish a baseline against which proposed improvements can be compared.

One danger in measuring a process is that what you measure becomes the priority, and some ways of improving one parameter may deteriorate other critical parameters. Optimizing a process may therefore adversely impact a larger process in which it is imbedded, or the quality of the process by other measures. For example, improvements in the cycle time for proposal evaluations can be made by taking less care in doing the work, for a loss in quality measured by the number of errors. As always, the test of value added is the overall impact on the customer. (Chapter 5 will discuss this further).

This chapter's summary matrix:

MANUFACTURING-BASED QUALITY MEASURES OF KNOWLEDGE WORK

PURPOSE:	RATE CUSTOMER SATISFACTION	APPRAISE AGENCY	IMPROVE PRODUCTS AND PROCESSES
MEASURE:	Program or Product line: Defect rates Cp or Cpk Cycle times	Aggregates of: Defect rates Cp or Cpk Cycle times	Process parameters: Defect rates Cp or Cpk Cycle times
TYPE OF MEASURE:	Statistical	Statistical	Statistical

5. Value-based Quality Measures

In value-based quality definitions, cost is a consideration. A low cost car which provides dependable and reasonably comfortable transportation would be considered a quality vehicle even if it does not have the features of a Rolls-Royce. In fact, the Rolls-Royce may be considered too expensive for what it provides and hence not good value for the average consumer. An Aerospace analogy may be the question of whether it is better to have a few expensive hi-tech fighters or a lot of cheaper, less capable models. Hence, measures of quality are not independent of cost.

Quality is also not independent of schedule. As discussed in Chapter 4, cycle time is a measure of quality, but improving cycle time can adversely affect other facets of quality such as defect rates. Conversely, a good product delivered late may be of no use to the customer. This is probably more often true of the products of knowledge work than those of assembly lines.

The author's view of value-based quality is that every product, service or process can be measured in three dimensions: cost, time, and some measure of "goodness," such as percent defects. Improvements which change one without detriment to the other two are always worthwhile. Other changes may or may not be worthwhile depending on the overall effect on the customer. While the trade-offs between cost, schedule and "goodness" can be a subjective matter, all quality decisions should try to balance the three considerations. For example, contracting can be measured by cycle time (schedule), overhead man-hours (cost) and number of protests per contract (defects). Measuring only one of these invites sacrificing the others.

For ease of reference, let us call a balanced combination of cost, schedule and "goodness" measurements a "quality troika." To illustrate the importance of balancing the three considerations, let us consider the problem of poor quality parts, which can be attacked by improving the part manufacturing process or by culling out defective parts through inspection. The former will reduce defects, lower costs and possibly shorten delivery time, while the latter will improve quality with attendant increases in costs and delays in delivery. Yet, both solutions have been used in actual cases.

Note that we measure cost and schedule constantly. We must do the same for the third dimension of Quality if we want credibility. Also if we reward cost/schedule adherence, we had better reward (which means measure) the other dimension of Quality.

In dealing with costs, we must recognize the difference between immediate and life cycle costs and that saving producer costs at the expense of customer costs may backfire. For example, we reduce the effort to verify a computer program which results in bugs in the customers application. Our costs are less, but the customer is unhappy. Schedule can also be sub-optimized, for example by shortening planning efforts which results in a longer execution effort.

Another approach to using value-based measures is to distinguish between effectiveness and efficiency. Effectiveness measures the "goodness" of a product or service for its user, while efficiency considers the cost of making it happen. To illustrate the difference, consider again the example of supplying integrated circuits meeting the customers needs by making much more than ordered and culling the output. We may wind up with enough quality products; we will wind up with a lot of scrap. Our customer may be pleased with the product (we are effective), but the cost of quality will be higher than it should be (we are not efficient). Effectiveness can be measured perhaps by sales (or the laboratory equivalent: amount of external funding), market share, or one of the product-based measures. Efficiency is measured by the cost of quality, overhead rates, or one of the manufacturing-based measures.

The Cost of quality is another concept developed by Crosby (6). It includes the cost of preventing defects, the cost of inspection, the cost of rework and the cost of waste. Unfortunately, as Deming notes (2), it also includes immeasurable costs, such as the cost of a lost customer. Many companies look only at the first two costs, considering only the money spent by their quality professionals (in prevention and inspection) as the cost of quality. In reality, a typical company may be spending 25% of their manufacturing costs on rework and scrap. This is a cost of quality. The cost of quality includes the cost of doing things wrong as well as the costs of preventing defects. For example, trying to save money by buying low grade IC's may result in a cost of rework far exceeding the price difference in parts. As Norman Augustine so aptly put it: it costs a lot to build bad products. (7)

It is an axiom of TQM that more effort in preventing defects is repaid many times over in savings in the other cost areas for an overall lower cost of quality. One way of measuring quality, from the standpoint of efficiency, could therefore be the determination of the measurable components of the cost of quality. The lower the cost of quality, the higher the efficiency of the quality effort.

Still another approach is the Taguchi loss function, which considers any product not meeting the design center to be of lesser quality as a function of its variation, even though it may still be within the specification limits (8). There are actually several loss functions, covering the cases where the product has a target value, where bigger is better, and where smaller is better. In all cases the calculated loss increases with the square of the deviation from the target. The loss can represent actual costs for repair of a defect, lost business, etc., or intangible losses such as the "loss to society" because of poor quality.

One way to matrix this chapter's information would be:

TO MEASURE:	EFFECTIVENESS	EFFICIENCY	COMBINATIONS
USE:	Sales Market share	Cost of quality Overhead rates Loss functions	quality troikas

Making it compatible with the matrixes in the other chapters:

VALUE-BASED QUALITY MEASURES OF KNOWLEDGE WORK

PURPOSE:	APPRAISE AGENCY	IMPROVE PRODUCTS AND PROCESSES
MEASURE:	Cost of Quality Overhead rates sales, market share	Quality troikas Loss functions
TYPE OF MEASURE:	Financial	Hybrid

6. User-based Quality Measures

As stated in chapter 1, all measures of quality must ultimately be user-based. The problem is translating user satisfaction to an appropriate quality measure. The most quoted user-based definition of quality is that of J. M. Juran (9), who defined quality as fitness for use.

Juran divides fitness for use into two categories: features and freedom from deficiencies. Features, he stated, cost money and attract customers, while freedom from defects saves money and keeps customers. Knowledge work features could include innovations, responsiveness, ease of comprehension of ideas presented, etc. and freedom from defects includes accuracy, legibility of written reports, etc.

Under this definition, product-based quality measures become user-based measures for evaluating features and manufacturing-based measures become user-based measures for evaluating freedom from defects. Transcendent and value-based quality measures may measure either features, freedom from defects, or overall fitness for use, depending on application

Using Juran's definition of quality as the starting point, the various measures separate (roughly) as shown in the following matrix:

TO MEASURE:	FEATURES	FREEDOM FROM DEFECTS	OVERALL FITNESS
MEASURE:	Rating scales	Defect rates	Climate indicators
	Product parameters	Cp or Cpk	Sales
	Performance indexes	Cycle times	Market share
	Systems effectiveness	Cost of quality	Quality troikas
		Overhead rates	
		Loss functions	

Recombining these into the usual format provides a matrix summarizing all the discussion so far:

USER-BASED QUALITY MEASURES FOR KNOWLEDGE WORK

PURPOSE:	RATE CUSTOMER SATISFACTION	APPRAISE AGENCY	APPRAISE INDIVIDUALS	IMPROVE PRODUCTS AND PROCESSES
MEASURE:	Rating scales Product parameters Performance indexes Systems effectiveness Defect rates Cp or Cpk Cycle times	Rating scales Climate indicators Defect rates Cp or Cpk Cycle times Cost of quality Overhead rates Sales, Market share	Rating scales	Defect rates Cp or Cpk Cycle times Quality troikas Loss functions
TYPE OF MEASURE:	Subjective, Objective, or Statistical.	Subjective, Surrogate, Statistical, or Financial.	Subjective.	Statistical or Hybrid.
DEFINITION OF QUALITY:	Transcendent, Product-based, or Manufacturing-based.	Transcendent, Product-based, Manufacturing-based, or Value-based.	Transcendent.	Manufacturing-based or Value-based.

7. Recommendations

Far from having no measures of the quality of knowledge work, we seem to have a plethora of choices. Some priority should therefore be established to guide a laboratory's approach.

Although all the considerations discussed are important, I believe a laboratory should put effectiveness before efficiency (see Chapter 5 for discussion) and features before freedom from defects (see Chapter 6). Effectiveness makes a potential customer interested in the work. Efficiency makes the purchase more affordable, but the interest must be there before this is relevant. Features are more important than freedom from defects for similar reasons. For a laboratory, producing "high tech" inefficiently is preferable to producing low tech efficiently. Of course, producing high tech efficiently is the best of all. Reducing the cost of quality is equivalent to finding more funds. More importantly, being both effective and efficient may possibly be the only way to survive.

The customer's transcendent evaluation of your quality is a subjective measure of considerable import. If your agency fails that test, all previous positive measures become meaningless. Hence, it is the first measure that should be obtained, if possible. Needed next are efficiency measures to assure you are competitive and help you remain so. All critical processes, such as contracting, should be measured for continual improvement of those things within the control of the agency which contribute to the customer's opinion of its quality or to the affordability of its products. Specific products and programs should have appropriate quality measures developed by the appropriate managers in the laboratory, working with their specific customers. This adds up to a lot of measurement, but if a product, program or process is important, it calls for a quality measure. In addition, you will need to establish benchmarks to compare against your measurements.

The author therefore recommends:

1. Survey your customers. Are you meeting their needs? If not, get their feelings on what needs attention. While the rating scales discussed in Chapter 2 are good for periodic surveys, I suggest the first survey be a face-to-face conference. The insights gained will be priceless.

2. With the aid of senior staff and, if available, outside peers, identify an appropriate set of surrogate measures to monitor for an evaluation of overall effectiveness (see Chapter 3). Rate other agencies to establish benchmarks and set goals for improvement. While you are at it, review your appraisal system, and set up subjective measures, of things that really count, for individual performance (see Chapter 2).

3. With the help of your workers, identify an appropriate set of measures of overall efficiency (see Chapter 5), and quality measures for all critical processes (see Chapter 4). Create appropriate process action teams to improve the processes, which will, in turn, enable improvements in the products and programs they service.

4. Work with individual customers to identify product or program measures which balance cost, schedule and "goodness" to the satisfaction of the customer (see Chapter 5). Set up measurement systems to identify problems and aid in constant improvement of product lines and program services.

5. Periodically review the operations of your quality measurement systems. Look for gaming problems, sub-optimizations, data availability, statistical analysis problems. Keep data collection as simple as possible. Quality measurement, too, is a process which should be constantly improved.

6. Most importantly, establish an atmosphere of cooperation and trust and make constant improvement a common goal for all employees from the commander to the lowest ranking. In such an environment, measurements of quality need not be imposed; they will spring up spontaneously.

As a final summary: There are valid ways to measure quality in the laboratory environment. It is decidedly not easy, but the alternative is to bet your future without knowing where you stand.

References

1. Garvin, A. David, "Managing Quality," The Free Press, NY, 1988

2. Deming, W. Edwards, "Out of the Crisis," MIT, MA, 1986

3. Weapon System Effectiveness Industry Advisory Committee, "Chairman's Final Report," AFSC-TR-65-6, Air Force Systems Command, 1965. AD-467816

4. Gunning, Robert, "New Guide to More Effective Writing in Business and Industry," Industrial Education Institute, Boston, 1964

5. Prichard, et al, University of Houston, "Organization Productivity Measurement: The Development and Evaluation of an Integrated Approach," AFHRL-TR-86-64, Air Force Human Resources Laboratory, 1987. AD-A183 565.

6. Crosby, Philip B., "Quality is Free," McGraw-Hill Book Co., NY, 1979

7. Augustine, Norman R., "Augustine's Laws" Penguin Books, NY, 1987

8. Ross, Phillip J., "Taguchi Techniques for Quality Engineering," McGraw-Hill Book Co., NY, 1988

9. Juran, J. M., "Juran on Leadership for Quality," The Free Press, NY, 1989

MISSION

OF

ROME LABORATORY

Rome Laboratory plans and executes an interdisciplinary program in research, development, test, and technology transition in support of Air Force Command, Control, Communications and Intelligence (C^3I) activities for all Air Force platforms. It also executes selected acquisition programs in several areas of expertise. Technical and engineering support within areas of competence is provided to ESD Program Offices (POs) and other ESD elements to perform effective acquisition of C^3I systems. In addition, Rome Laboratory's technology supports other AFSC Product Divisions, the Air Force user community, and other DOD and non-DOD agencies. Rome Laboratory maintains technical competence and research programs in areas including, but not limited to, communications, command and control, battle management, intelligence information processing, computational sciences and software producibility, wide area surveillance/sensors, signal processing, solid state sciences, photonics, electromagnetic technology, superconductivity, and electronic reliability/maintainability and testability.